

1 **Improvement of short-term forecasting**
2 **in the Northwest Pacific through assimilating Argo**
3 **data into initial fields**

4
5 FU Hongli¹, Peter C. Chu², HAN Guijun^{1*}, HE Zhongjie¹, LI Wei¹, ZHANG Xuefeng¹

6 1 Key Laboratory of Marine Environmental Information Technology, State Oceanic
7 Administration, National Marine Data and Information Service, Tianjin 300171, China

8 2 Naval Ocean Analysis and Prediction (NOAP) Laboratory

9 Naval Postgraduate School, Monterey, CA, USA

10
11
12
13
14
15
16

*Corresponding author

Address: 93 Liuwei Road, Hedong District, Tianjin 300171, China.

Email: fhlkjj@163.com

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 2012		2. REPORT TYPE		3. DATES COVERED 00-00-2012 to 00-00-2012	
4. TITLE AND SUBTITLE Improvement of short-term forecasting in the Northwest Pacific through assimilating Argo data into initial fields				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School,Naval Ocean Analysis and Prediction (NOAP) Laboratory,Monterey,CA,93943				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES Acta Oceanologica Sinica, Chinese Society of Oceanography, in press					
14. ABSTRACT The impact of assimilating Argo data into initial field on the short-term forecasting accuracy of temperature and salinity is quantitatively estimated by using a forecasting system of the western North Pacific, on the base of the Princeton Ocean Model with generalized coordinate system (POMgcs). This system uses a sequential multi-grid three-dimensional variational (3DVAR) analysis scheme to assimilate observation data. Two numerical experiments were conducted with and without Argo temperature and salinity profile data besides conventional temperature and salinity profile data and sea surface height anomaly (SSH_a) and sea surface temperature (SST) in the process of assimilating data into initial fields. The forecast errors are estimated through using independent temperature and salinity profiles during the forecasting period, including the vertical distributions of the horizontally averaged root mean square errors (H-RMSEs) and horizontal distributions of the vertically averaged mean errors (MEs) and temporal variation of spatially averaged root mean square errors (S-RMSEs). Comparison between the two experiments shows that the assimilation of Argo data significantly improves the forecast accuracy, with 24% reduction of H-RMSE maximum for the temperature, and the salinity forecasts are improved more obviously, averagely dropping of 50% for H-RMSEs in depth shallower than 300m. Such improvement is caused by relatively uniform sampling of both temperature and salinity from the Argo drifters in time and space.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 31	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

17 **Abstract**

18 The impact of assimilating Argo data into initial field on the short-term
19 forecasting accuracy of temperature and salinity is quantitatively estimated by using a
20 forecasting system of the western North Pacific, on the base of the Princeton Ocean
21 Model with generalized coordinate system (POMgcs). This system uses a sequential
22 multi-grid three-dimensional variational (3DVAR) analysis scheme to assimilate
23 observation data. Two numerical experiments were conducted with and without Argo
24 temperature and salinity profile data besides conventional temperature and salinity
25 profile data and sea surface height anomaly (SSHa) and sea surface temperature (SST)
26 in the process of assimilating data into initial fields. The forecast errors are estimated
27 through using independent temperature and salinity profiles during the forecasting
28 period, including the vertical distributions of the [horizontally averaged root mean](#)
29 [square errors \(H-RMSEs\)](#) and horizontal distributions of the vertically averaged mean
30 errors (MEs) and temporal variation of spatially averaged [root mean square errors](#)
31 [\(S-RMSEs\)](#). Comparison between the two experiments shows that the assimilation of
32 Argo data significantly improves the forecast accuracy, with 24% reduction of
33 [H-RMSE](#) maximum for the temperature, and the salinity forecasts are improved more
34 obviously, averagely dropping of 50% for [H-RMSEs](#) in depth shallower than 300m.
35 Such improvement is caused by relatively uniform sampling of [both temperature and](#)
36 [salinity from the Argo drifters](#) in time and space.

37 Key words: Data assimilation, Argo data, Western North Pacific, Ocean prediction

38 1. Introduction

39 Data assimilation, required in operational ocean data retrieval, has contributed
40 significantly to the success of ocean prediction. It is to blend modeled variable (x_m)
41 with observational data (y_o) (Chu et al., 2004; Chu et al., 2010),

$$42 \quad x_a = x_m + W \bullet [y_o - H(x_m)] \quad (1)$$

43 where x_a is the assimilated variable; H is an operator that provides the model's
44 theoretical estimate of what is observed at the observational points, and W is the
45 weight matrix. Difference among various data assimilation schemes such as optimal
46 interpolation (Chu et al., 2007a; Chu et al., 2007b), Kalman filter (Galanis et al.,
47 2011), and three-dimensional variational (3DVAR) methods (Li et al., 2008) is the
48 different ways to determine the weight matrix W. The data assimilation process (1)
49 can be considered as the average (in a generalized sense) of x_m and y_o . The two parts
50 (x_m and y_o) in the assimilation process usually have very different characteristics in
51 terms of data temporal and spatial distribution: uniform and dense in the modeled data
52 (x_m), and non-uniform and sparse in the observed data (y_o). Question arises: What is
53 the impact of data sampling strategies in the assimilation of initial field on the
54 forecasting accuracy? To answer this question, two observational datasets are needed
55 with different types of data distribution patterns in space and time. One is relatively
56 uniform, and the other is not.

57 The Global Temperature and Salinity Profile Program (GTSP), as a cooperative
58 international project, has been established since 1990 to provide global temperature (T)

59 and salinity (S) resources. GTSPP contains conventional temperature and salinity
60 profile data such as Nansen bottle, conductivity-temperature-depth (CTD), and
61 bathythermograph (BT), which are usually collected from ships. Since the Array for
62 Real-time Geostrophic Oceanography (Argo) is launched into practice, GTSPP (T, S)
63 profiles increase rapidly in both quantity and quality. It becomes possible to monitor
64 the temporal and spatial variations of temperature and salinity simultaneously. Liu et
65 al. (2004) showed significant improvement of temperature prediction in the central
66 Pacific using a global ocean model with Argo data assimilation. Griffa et al. (2006)
67 analyzed the impact of Argo data assimilation on a Mediterranean prediction model
68 by a set of idealized experiments, and discussed the impact of coverage density and
69 locations of Argo data on assimilation results.

70 Due to the limitation of ship time, the conventional (T, S) profile data are
71 non-uniformly distributed in space and time. However, the Argo floats drift freely
72 with ocean currents, the Argo data are more uniformly distributed in space and time
73 than the conventional data. Such difference in data distributions between the
74 conventional (non-uniform) and Argo (relatively uniform) (T, S) profile data provides
75 an opportunity to study the effect of the sampling strategies on the ocean prediction
76 accuracy. To do so, a numerical forecasting system with 3DVAR in the western
77 Pacific regional seas (Fig. 1) is constructed with the capability to assimilate sea
78 surface height anomaly (SSHa) from altimeters and sea surface temperature (SST)
79 from satellite remote sensors, as well as in-situ conventional and Argo (T, S) profiles
80 in the determining of the initial conditions. A seven-day forecast is conducted with

81 and without the assimilation of Argo (T, S) profiles in initial field. The prediction
82 accuracy is verified with independent temperature and salinity profiles during the
83 period of prediction (not used in the data assimilation of initial field). Difference
84 between the two forecast experiments shows the impact of data distribution on the
85 ocean prediction accuracy.

86 Frame of the paper is outlined as follows. Section 2 shows the basic features of
87 conventional and Argo profile data. Section 3 describes the ocean dynamic model and
88 ocean data assimilation scheme. Section 4 gives the experiment design and the
89 quantitative analysis on the improvement of ocean prediction using the Argo data
90 assimilation. Section 5 presents the conclusions.

Figure 1

92 **2. Data**

93 Ocean observational Data (January-December 2008) include SSHa from
94 multi-satellite altimeters and SST from satellite remote sensors, and (T, S) profiles
95 (conventional and Argo) from GTSP. The satellite SSHa and SST data are on the
96 horizontal resolution of 0.25° and the time increment of 1 day. Quality control is
97 conducted on both conventional and Argo profile data before assimilating them into
98 the initial field of the numerical forecasting. For the conventional data, it includes
99 position/time check, depth duplication check, depth inversion check, temperature and
100 salinity range check, excessive gradient check, and stratification stability check. For

101 the Argo floats, it includes duplicate float test, land position test, float drafting
102 velocity test, pressure range test, temperature and salinity coherence test, pressure
103 level duplication test and pressure inversion test, spike test, salinity and temperature
104 gradient test, and stratification stability test, etc. In addition, the calibration method
105 developed by Wong et al. (2003) is employed to calibrate the sensor drift of salinity
106 measurements in the Argo data.

107 Figure 2 shows the horizontal distribution of (T, S) profile data. From January to
108 December 2008, there are 60634 temperature profiles and 52638 salinity profiles from
109 conventional observations, 5323 temperature profiles and 5210 salinity profiles from
110 Argo floats. That is to say, the Argo data is near 1/10 of the conventional data. The
111 conventional (T, S) profiles are distributed non-uniformly in horizontal with most
112 profiles around Japan and east of Taiwan and much less profiles in the other regions,
113 and existence of some data-void areas. The Argo (T, S) profiles are distributed
114 uniformly (relative) over the whole area. Figure 3 shows the vertical distributions of
115 numbers of observations for temperature and salinity from conventional and Argo data.
116 The conventional temperature (salinity) observations decrease slowly from 57597
117 (48595) data points near the surface to about 40000 (T and S) data points at near 700
118 m depth, and reduce drastically to around 2000 (T and S) data points below 700 m
119 depth (Fig. 3a). The Argo temperature (salinity) observations have 5299 (5186) data
120 points from near surface to about 420 m depth, decrease almost linearly to 2000 (T
121 and S) data points at about 1500 m depth, keep 2000 (T and S) data points from 1500
122 to 1800 m depth, and reduce to less than 100 data points at 2000 m depth (Fig. 3b).

123 Two (T, S) datasets are used to investigate the impact of the sampling strategies
124 on the ocean prediction accuracy. The first dataset (called “WITH_ARGO”) contains
125 Argo profile data besides conventional profiles, SSHa and SST and represents
126 horizontally uniform (relative) sampling. The second dataset (called “NO_ARGO”)
127 contains only the conventional profile data, SSHa and SST and represents horizontally
128 non-uniform sampling.

129  Figures 2, 3

130 **3. Ocean Prediction System**

131 **3.1 Ocean Model**

132 The ocean model used in this study is the Princeton Ocean Model with
133 generalized coordinate system (POMgcs). The study domain covers from 99°E to
134 150°E in longitude, and from 10°N to 52°N in latitude (Fig. 1), with variable
135 horizontal resolution starting from 1/12° near the coastal waters of China and
136 Kuroshio, and telescoping to 1/2° at other areas. The vertical coordinate is a
137 combination of sigma and z-level with a maximum depth of 5035 m, discretized by 35
138 model levels. In the vicinity of upper mixed layer and thermocline, z-coordinate is
139 adopted in order to get a higher vertical resolution. In shallow water and the area near
140 bottom boundary, the terrain-following σ -coordinate is used. Sea surface forcing
141 fields consist of winds, air temperatures, humidity and clouds from the National
142 Centers for Environmental Prediction (NCEP) reanalysis. Sea surface heat fluxes are

143 calculated by bulk formula, and open boundary conditions are provided by the
144 simulation results of Massachusetts Institute of Technology general circulation model
145 (MITgcm, Marshall et al., 1997), including daily Sea level, temperature, salinity, and
146 currents. These open boundary data are interpolated to the grid and time step of the
147 forecasting system.

148 **3.2 Ocean Data Assimilation Scheme**

149 The ocean data assimilation scheme used in the system is a sequential
150 three-dimensional variational (3DVAR) analysis scheme designed to assimilate
151 temperature and salinity using a multi-grid framework (Li et al., 2008). This
152 sequential 3DVAR analysis scheme can be performed in three dimensional spaces and
153 can retrieve resolvable information from longer to shorter wavelengths for a given
154 observation network and yield multi-scale analysis. The basic idea of this data
155 assimilation scheme can be referred to Li et al. (2008) and Li et al. (2010).

156 The data assimilation is carried out in the upper 1000m. The basic idea proposed
157 by Troccoli et al. (2002) is employed to make salinity adjustment for the background
158 field after temperature data is assimilated. The area extent of adjustment is limited
159 between the latitude of 30°S-30°N and depth of 50-1000m. It needs firstly to establish
160 a T-S relationship by using interpolation algorithm based on the instant model T-S
161 table. Then the background field of salinity is adjusted based on the T-S relationship
162 and temperature analysis result. In addition, an idea of converting satellite altimeter
163 SSHa into T-S “pseudo profiles” based on the 3DVAR scheme is adapted ((Zhu and

164 Yan, 2006; He et al., 2010).

165 Figure 4 shows the flow chart for data assimilation procedure: (1) Based on 24-h
166 forecasting (T, S) values, obtain the T-S relationship at every grid point through using
167 the T-S relationship module; (2) Convert altimeter SSHa into “pseudo profiles” of
168 temperature and salinity; (3) Assimilate temperature data to obtain temperature
169 analysis field; (4) Adjust 24-h forecasting salinity field on the base of the T-S
170 relationship and temperature analysis result, and take the adjusted salinity field as the
171 background field for salinity assimilation; (5) Assimilate salinity data to obtain
172 salinity analysis field; (6) the temperature and salinity analysis fields are used as the
173 initial conditions of next seven-day forecast.

Figure 4

175 3.3. Experiment Design

176 Two forecast experiments are designed. The first experiment (called
177 “NO_ARGO”) assimilates all available observations (conventional *T*, *S* profiles and
178 SSHa and SST) except the Argo profile data. The second experiment (called
179 “WITH_ARGO”) assimilates all available observations including the Argo profile
180 data. Both experiments use the same sea-surface forcing fields and open boundary
181 conditions. The China Ocean ReAnalysis (CORA) fields of January 1, 2008 (Han et
182 al., 2011, <http://www.cora.net.cn>) are used as initial conditions. First, a seven-day
183 forecast is performed for both experiments. Second, the data assimilation is performed
184 using 24-hour forecast values as the background field. Taking the assimilated fields as

185 initial conditions, the next seven-day forecast is performed. This procedure
186 (forecast-assimilation-forecast) is cycled 365 times to obtain 24-hour, 48-hour,
187 72-hour, 96-hour, 120-hour, 144-hour, 168-hour forecast values of temperature and
188 salinity fields in every day of 2008. The time window of assimilating SST and SSHa
189 data in both experiments is set to one day, namely assimilating satellite data within the
190 one day before initial forecasting time. Since the spatial distributions of conventional
191 observations and Argo data are sparse, both experiments adopt the 3.5-day time
192 window, namely assimilating ocean (T , S) profile data within the 3.5 days before
193 initial forecasting time. Since all temperature and salinity observational data during
194 the period of forecasting are not assimilated into background fields (initial field of the
195 numerical forecasting), they are taken as independent data to be used to check the
196 forecast result. Based on these independent observation data, the errors of the 24-hour,
197 48-hour, 72-hour, 96-hour, 120-hour, 144-hour, and 168-hour forecast values of the
198 temperature and salinity at each grid point in every day of 2008 can be estimated. The
199 vertical distributions of forecast errors are obtained by averaging the errors in the
200 horizontal direction. The horizontal distributions of forecast errors are obtained by
201 averaging the errors in the vertical direction. Difference of forecast errors between the
202 two experiments shows the effect of sampling strategies on the ocean prediction
203 accuracy.

204 4. Effect of Argo Data

205 4.1 Whole 3D Domain

206 To quantify the impact of assimilating Argo data on an ocean prediction errors,
207 the horizontally averaged root mean square error (H-RMSE) between predicted and
208 observed values for the whole horizontal region at depth z_k and time t_m is calculated
209 by

$$210 \quad \text{H-RMSE}^{(\psi)}(z_k, t_m) = \sqrt{\frac{1}{N} \sum_{n=1}^N [\psi^p(x_n, y_n, z_k, t_m) - \psi^o(x_n, y_n, z_k, t_m)]^2} \quad (2)$$

211 where x_n and y_n indicate the zonal and latitudinal coordinates of the n th observation
212 point, respectively; z_k is the depth of the k th level; t_m is the m th forecasting time; N is
213 total number of observation points at the t_m time and z_k depth; $\psi^p(x_n, y_n, z_k, t_m)$ and
214 $\psi^o(x_n, y_n, z_k, t_m)$ respectively denote the predicted and ground-truth values at the t_m
215 time and z_k depth for the point (x_n, y_n) . In the study, ψ indicates temperature (T) or
216 salinity (S). $\text{H-RMSE}^{(\psi)}(z_k, t_m)$ can be used to evaluate the overall performance for
217 the whole depths.

218 Figure 5 a and b show the vertical distribution of $\text{H-RMSEs}^{(T)}$ for $t_1=24$ -hour and
219 $t_2=168$ -hour forecasts with and without Argo profiles assimilation. Since the high
220 resolution and horizontally uniform satellite remote sensing SST data are assimilated,
221 inclusion of Argo data does not improve the accuracy of SST prediction.

222 $\text{H-RMSEs}^{(T)}$ at time t_1 and t_2 increase with depth from the surface to its
223 maximum value at around 158 m depth, where is the mean thermocline location,
224 reduce drastically to 0.5°C at around 1000 m depth, and reduce gradually to 0.25°C

225 to 2000 m depth. The low value of $H\text{-RMSE}^{(T)}$ below 1000 m depth for all cases may
226 be caused by the low variability.

227 For 24-hour forecast (Fig. 5a), the maximum value of $H\text{-RMSE}^{(T)}$ is 2.1°C
228 without Argo data assimilation and 1.6°C with Argo data assimilation (24% error
229 reduction). The improvement of ocean prediction is very evident until 1000 m depth.
230 Since the value of $H\text{-RMSE}^{(T)}$ below 1000 m depth is already small ($0.25\text{--}0.5^{\circ}\text{C}$), the
231 improvement with the Argo data is not noticeable. Such improvement in upper 1000
232 m especially at around 158 m depth is still evident in 168-hour forecast (Fig. 5b).

Figure 5

233
234 Figure 5 c and d show the vertical distribution of $H\text{-RMSEs}^{(S)}$ for $t_1=24\text{-hour}$ and
235 $t_2=168\text{-hour}$ forecasts with and without Argo profile data assimilation. Similar to the
236 temperature prediction, the $H\text{-RMSE}$ of salinity for all cases reduces evidently from
237 the surface to depth around 1200 m, and reduces gradually below 1200 m. The low
238 value of $H\text{-RMSE}^{(S)}$ below 1200 m depth is related to the low variability. Without
239 Argo data assimilation, $H\text{-RMSEs}^{(S)}$ at time t_1 and t_2 are very large, with more than
240 0.5 psu for depths shallower than 300 m. With Argo data assimilation, they decrease
241 drastically to less than 0.23 psu for 24-hours forecast and 0.25 psu for 168-hour
242 forecast with error reduction more than 50%. Below 1200 m depth, $H\text{-RMSEs}^{(S)}$ at
243 time t_1 and t_2 are quite small with slightly larger values in “WITH_ARGO”
244 experiment than in the “NO_ARGO” experiment. This may be related that the depth
245 of assimilating data is limited to upper 1000m. A further study is needed to explain
246 such phenomena.

247 4.2 Near Thermocline

248 The mean errors (ME) within the layers between z_{k1} and z_{k2} at time t_m is
249 calculated using Eq.(3) to identify the forecast system performance.

$$250 \quad ME_{k_1, k_2}^{(\psi)}(x_n, y_n, t_m) = \frac{1}{K} \sum_{k=k_1}^{k_2} (\psi^p(x_n, y_n, z_k, t_m) - \psi^o(x_n, y_n, z_k, t_m)) \quad (3)$$

251 Where all letters express the same means as ones in the Eq.(2) and k_1, k_2 represents
252 the k_1 th and k_2 th level, respectively; K equals to $k_2 - k_1 + 1$. Here, to evaluate the forecast
253 performance near the mean thermocline, the depths of the k_1 th and k_2 th level are 100m
254 and 300m, respectively, and the t_m is 24-hour.

255 Figure 6 a and b show the horizontal distributions of the vertically (100-300 m)
256 averaged temperature mean errors in 24-hour forecast without and with Argo data
257 assimilation, respectively. Without Argo data assimilation, the predicted temperatures
258 are lower than observations in most areas. In the east areas of Japan, the predicted
259 temperatures are 0.8°C higher than observations. With Argo data assimilation, the
260 predicted temperatures are significantly improved, and the forecast errors are 0.1°C
261 or less in the whole areas. Therefore, the assimilation of Argo data can reduce errors
262 of temperature forecast dramatically near the mean thermocline.

Figure 6

263
264 Figure 6 c and d show the horizontal distributions of the vertically (100-300 m)
265 averaged salinity mean errors in 24-hour forecast without and with Argo data
266 assimilation, respectively. Without Argo data assimilation, the predicted salinity is
267 significantly lower than observations in most areas. For example, the predicted

268 salinity is over 0.5 psu lower than observation in the area of 15°N-35°N. However, the
 269 predicted salinity is significantly higher than observation in the small east area of
 270 Japan. It indicates that an obvious bias exists for salinity forecast without Argo data
 271 assimilation. With Argo data assimilation, the predicted salinity is significantly
 272 improved, and the forecast errors are 0.2 psu or less in the whole areas. Therefore, the
 273 assimilation of Argo data can reduce errors of salinity forecast dramatically near the
 274 mean halocline.

275 **4.3 Error Evolution**

276 The [spatially averaged root mean square error \(S-RMSE\)](#) between predicted and
 277 observed values for the whole horizontal region within the layers between z_{k1} and z_{k2}
 278 and at time t_m ,

$$279 \quad \text{S-RMSE}_{k_1, k_2}^{(\psi)}(t_m) = \sqrt{\frac{1}{NK} \sum_{k=k_1}^{k_2} \sum_{n=1}^N [\psi^p(x_n, y_n, z_k, t_m) - \psi^o(x_n, y_n, z_k, t_m)]^2} \quad (4)$$

280 is also used for the evaluation. Just as Eq.(3), all letters in the Eq.(4) express the same
 281 means as ones in the Eq.(2).

282 The [S-RMSEs](#) of temperature are calculated using Eq.(4) for upper (0–50m) and
 283 lower (50–1000m) layers to analysis the errors growth (Fig. 7). The [S-RMSEs^{\(T\)}](#) are
 284 generally lager and grow faster in the upper layer than in the lower layer. For the
 285 upper layer, without Argo data assimilation, the [S-RMSE^{\(T\)}](#) is 1.33°C for 24-hour
 286 forecast, and 1.51°C for 168-hour forecast (14% increasing). With Argo data
 287 assimilation, the [S-RMSE^{\(T\)}](#) is 1.26°C for 24-hour forecast, and 1.49°C for 168-hour
 288 forecast (18% increasing). For the lower layer, without Argo data assimilation, the

289 $S\text{-RMSE}^{(T)}$ is 1.15°C for 24-hour forecast, and 1.18°C for 168-hour forecast (3%
290 increasing). With Argo data assimilation, the $S\text{-RMSE}^{(T)}$ is 0.93°C for 24-hour
291 forecast, and 1.03°C for 168-hour forecast (11% increasing).

292 With Argo data assimilation, the accuracy of temperature forecasts is
293 significantly improved. However, it is worthy note that the forecast errors in the
294 “WITH_ARGO” experiment grow a little faster compared to those in the
295 “NO_ARGO” experiment. This is because the assimilation of Agro data just improves
296 the accuracy of initial conditions and can not correct the model systematic bias. As a
297 result, the forecast error around initial forecast time in the “WITH_ARGO”
298 experiment is mainly determined by the accuracy of initial conditions and much lower
299 than ones in the “NO_ARGO” experiment, and with the increase of the forecast time,
300 the forecast error is mainly affected by model systematic bias so that the forecast error
301 with assimilation of Argo data increases sharply.

Figure 7, 8

302
303 Same as the temperature, the $S\text{-RMSEs}$ of salinity are calculated using Eq.(4) for
304 upper (0–300m) and lower (300–1000m) layers to identify the errors growth (Fig. 8).
305 $S\text{-RMSEs}^{(S)}$ are generally lager in the upper layer than in the lower layer. For the
306 upper layer, without Argo data assimilation, the $S\text{-RMSE}^{(S)}$ is near 0.5 psu for the
307 whole prediction period. With Argo data assimilation, the $S\text{-RMSE}^{(S)}$ is 0.17 psu for
308 24-hour forecast, and 0.22 psu for 168-hour forecast, much less than 50% of that
309 without Argo data assimilation. For the lower layer, without Argo data assimilation,
310 the $S\text{-RMSE}^{(S)}$ is near 0.15 psu for the whole prediction period. With Argo data

311 assimilation, the $S\text{-RMSEs}^{(S)}$ are 0.07 psu and 0.09 psu for 72-hour and longer
312 forecast, and the $S\text{-RMSEs}^{(S)}$ reduce around 40% relative to that without Argo data
313 assimilation. So, with Argo data assimilation, the accuracy of salinity forecasts is
314 significantly improved.

315

Figures 9

316 **4.4 Vertical Cross Sections**

317 A set of CTD temperature measurements (not being used in the data assimilation)
318 is used for the evaluation. It was conducted on 23 February 2008 along 129°E south
319 of Japan. Figure 9a gives the distribution of observational temperatures for the 129°E
320 cross-section, while Fig. 9b and c show results of 24-hour forecast for both
321 experiments. Temperature field with Argo data assimilation is closer to observations
322 than that without Argo data assimilation.

323 The section along 38.5°E east of Japan during 8 May 2008 is used for illustration.
324 Figure 10a gives the distribution of observational salinity, while Fig. 10b and c show
325 results of 24-hour forecast for both experiments. Just as temperature section, salinity
326 field with Argo data assimilation is closer to observations than that without Argo data
327 assimilation.

328 **5. Conclusion**

329 A forecast system based on the Princeton Ocean Model with generalized
330 coordinate system (POMgcs) and sequential multi-grid 3DVAR analysis scheme is

331 developed for the western Pacific marginal seas to investigate the impact of sampling
332 strategies on the ocean prediction through using two (T , S) profile datasets. The first
333 dataset contains both conventional and Argo profile data (called “WITH_ARGO”)
334 and represents horizontally uniform (relative) sampling. The second dataset contains
335 only the conventional profile data (called “NO_ARGO”) and represents horizontally
336 non-uniform sampling.

337 Without Argo data assimilation (i.e., non-uniform sampling), temperature and
338 salinity forecast have obvious biases. Especially in the area of 15°N-35°N the
339 predicted temperature and salinity are obviously smaller than observations. With Argo
340 data assimilation, these biases are corrected. Based on the detailed comparison of
341 horizontally averaged root mean square error (H-RMES) between the two
342 experiments, it is known that the temperature H-RMSE maximum drops by 24% and
343 the salinity H-RMSEs in depth shallower than 300m drop averagely by 50% if the
344 Argo data is assimilated into initial fields, and the accuracy of salinity forecast is
345 improved more obviously than temperature forecast. With Argo data assimilation, the
346 temperature or salinity distribution along some vertical cross sections is nearer to
347 observations than that without Argo data assimilation. It indicates that the assimilation
348 of Argo data plays an important role in the process of constructing initial fields, and it
349 can significantly improves the temperature and salinity forecasts. It is worthy note that
350 although the forecast errors within assimilation depth (shallower than 1000m) can be
351 sharply reduced though assimilating Argo data into initial filed, the errors below
352 1000m depth change very small, or even can slightly increase. A further study is

353 needed to explain such phenomena.

354 **Acknowledgements**

355 This study was jointly supported by grants of the National Natural Science
356 Foundation of China (41030854, 40906015, 40906016, 41106005, and 41176003,
357 41206178), and National Science and Technology Support program
358 (2011BAC03B02-01-04).

359 **Reference**

- 360 Chu Peter C, Wang GuiHua, Fan Chenwu. 2004. Evaluation of the U.S. Navy's
361 Modular Ocean Data Assimilation System (MODAS) using the South China Sea
362 Monsoon Experiment (SCSMEX) data. *J. Oceanogr.*, **60**: 1007-1021
- 363 Chu Peter C, Amezaga G R, Gottshall E L, et al. 2007a. Ocean nowcast/forecast
364 systems for improvement of Naval undersea capabilities. *Marine Technol. Soc. J.*,
365 **41**(2): 23-30
- 366 Chu Peter C, Mancini S, Gottshall E L, et al. 2007b. Sensitivity of satellite altimetry
367 data assimilation on weapon acoustic preset using MODAS. *IEEE J. Oceanic*
368 *Eng.*, **32**: 453-468
- 369 Chu Peter C, Fan Chenwu. 2010. A conserved minimal adjustment scheme for
370 stabilization of hydrographic profiles. *J. Atmos. Oceanic Technol.*, **27**(6):
371 1072-1083
- 372 Galanis G, Chu Peter C, Kallos G. 2011. Statistical post processes for the

373 improvement of the results of numerical wave prediction models. A combination
 374 of Kolmogorov–Zurbenko and Kalman filters. *J. Operat. Oceanogr.*, **4**(1): 23-31
 375 Griffa A, Molcard A, Raicich F, et al. 2006. Assessment of the impact of TS
 376 assimilation from ARGO floats in the Mediterranean Sea. *Ocean Sci.*, **2**: 237-248
 377 Han Guijun, Li Wei, Zhang Xuefeng, et al. 2011. A regional ocean reanalysis system
 378 for China coastal waters and adjacent seas. *Advances in Atmospheric Sciences*,
 379 **28**(3): 682-690
 380 He Zhongjie, Han Guijun, Li Wei, et al. 2010. Experiments on assimilating of satellite
 381 data in the China seas and adjacent seas (in Chinese). *Periodical of Ocean*
 382 *University of China*, **40**(9): 1-7
 383 Li Wei, Xie Yuanfu, He Zhongjie, et al. 2008. Application of the multi-grid data
 384 assimilation scheme to the China Seas' temperature forecast. *J. Atmos. Oceanic*
 385 *Technol.*, **25**(11): 2106–2116
 386 Li Wei, Xie Yuanfu, Deng Shiuwming, et al. 2010. Application of the multigrid
 387 method to the two-dimensional doppler radar radial velocity data assimilation. *J.*
 388 *Atmos. Oceanic. Tech.*, **27**(2): 319-332
 389 Liu Yimin, Zhang Renhe, Yin Yonghong, et al. 2004. The Application of ARGO Data
 390 to the Global Ocean Data Assimilation Operational System of NCC. *Acta*
 391 *Meteorologica Sinica*, **19**: 355-365
 392 Marshall J, Hill C, Perelman L, et al. 1997. Hydrostatic, quasi-hydrostatic, and
 393 nonhydrostatic ocean modelling. *J. Geophys. Res.*, **102**(C3): 5733-5753
 394 Troccoli A, Balmaseda M A, Segschneider J, et al. 2002. Salinity adjustments in the

395 presence of temperature data assimilation. *Mon. Wea. Rev.*, **130**: 89-102

396 Wong A P S, Johnson G C, Owens W B. 2003. Delayed-mode calibration of

397 autonomous CTD profiling float salinity data by $S-\theta$ climatology. *J. Atmos.*

398 *Oceanic Tech.*, **20**:308-318

399 Zhu Jiang, Yan Changxiang. 2006. Nonlinear balance constraints in 3DVAR data

400 assimilation. *Science in China (D)*, **49**: 331-336

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416 **Figure**

417 Fig. 1. Geography of the Western North Pacific. The dots indicate the numerical grid
418 points.

419 Fig. 2. Spatial distribution of temperature (a) and salinity (b) profiles from GTSP
420 during Jan-Dec 2008 (Red dot: conventional data; Blue dot: Argo data).

421 Fig. 3. Vertical distributions of numbers of observations for temperature (red) and
422 salinity (blue) from conventional (a) and Argo data (b).

423 Fig. 4. Flow chart of multi-grid 3DVAR operational procedure.

424 Fig. 5. Vertical dependence of temperature (a, b, □) and salinity (c, d, psu) **H-RMSEs**
425 in 24-hour forecast (a, c) and 168-hour forecast (b, d) with and without Argo data
426 assimilation.

427 **Fig. 6.** Horizontal distribution of vertically (100-300 m) averaged temperature (a, b,
428 °C) and salinity (c, d, psu) prediction errors in 24-hour forecast without Argo profiles
429 assimilation(a, c) and with Argo profiles assimilation(b, d).

430 **Fig. 7.** Temporal variation of **temperature S-RMSEs** (°C) for the layers of 0-50m(a)
431 and 50-1000m(b) in 24-hour forecast with and without Argo data assimilation.

432 **Fig. 8.** Temporal variation of **salinity S-RMSEs** (psu) for the layers of 0-300m(a) and
433 300-1000m(b) in 24-hour forecast with and without Argo data assimilation.

434 **Fig. 9.** Vertical temperature cross-section along 129°E south of Japan on 23 February
435 2008: (a) observation (dark dots: stations), (b) 24-hour forecast without assimilating
436 Argo profiles, and (c) 24-hour forecast with assimilating Argo profiles.

437 **Fig. 10.** Vertical salinity cross-section along 38.5°N east of Japan on 8 May 2008: (a)
438 observation (dark dots: stations), (b) 24-hour forecast without assimilating Argo
439 profiles, and (c) 24-hour forecast with assimilating Argo profiles.

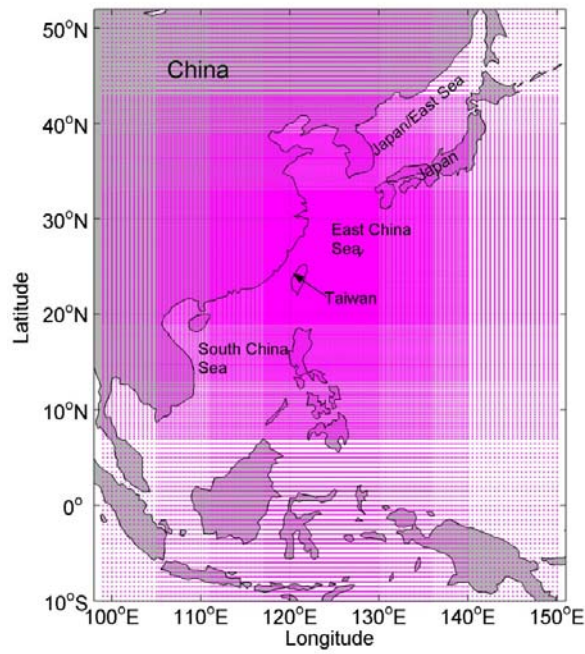


Fig. 1. Geography of the Western North Pacific. The dots indicate the numerical grid points.

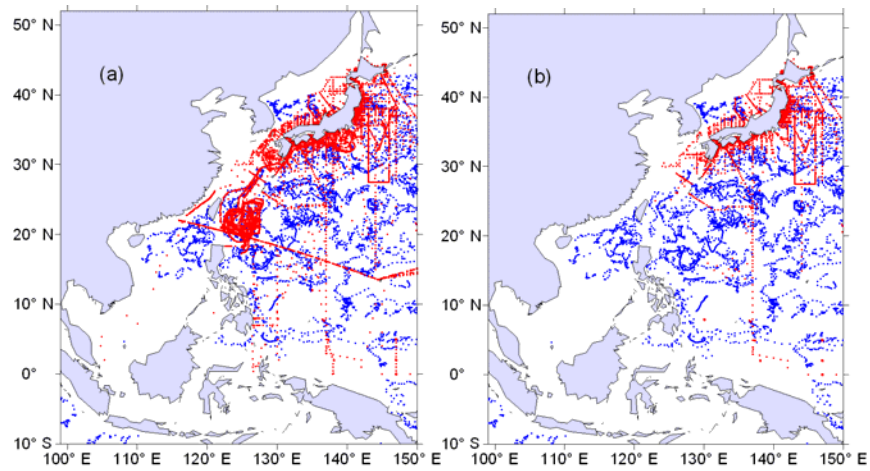


Fig. 2. Spatial distribution of temperature (a) and salinity (b) profiles from GTSP during Jan-Dec 2008 (Red dot: conventional data; Blue dot: Argo data).

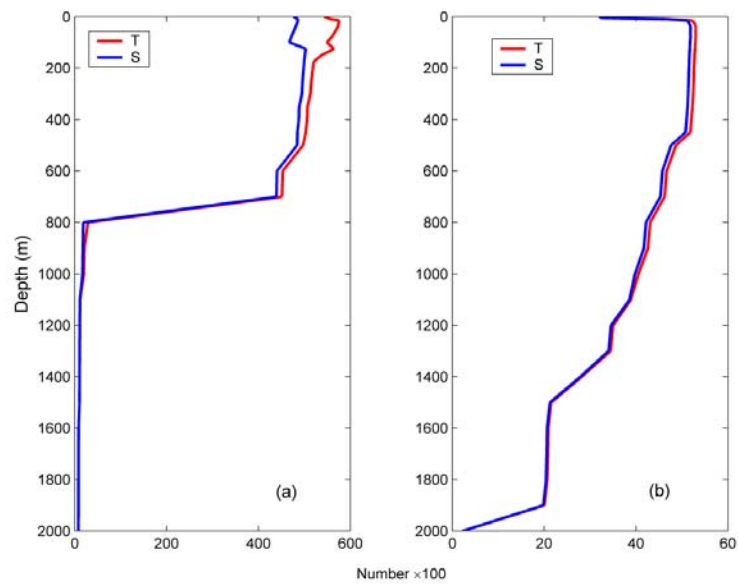
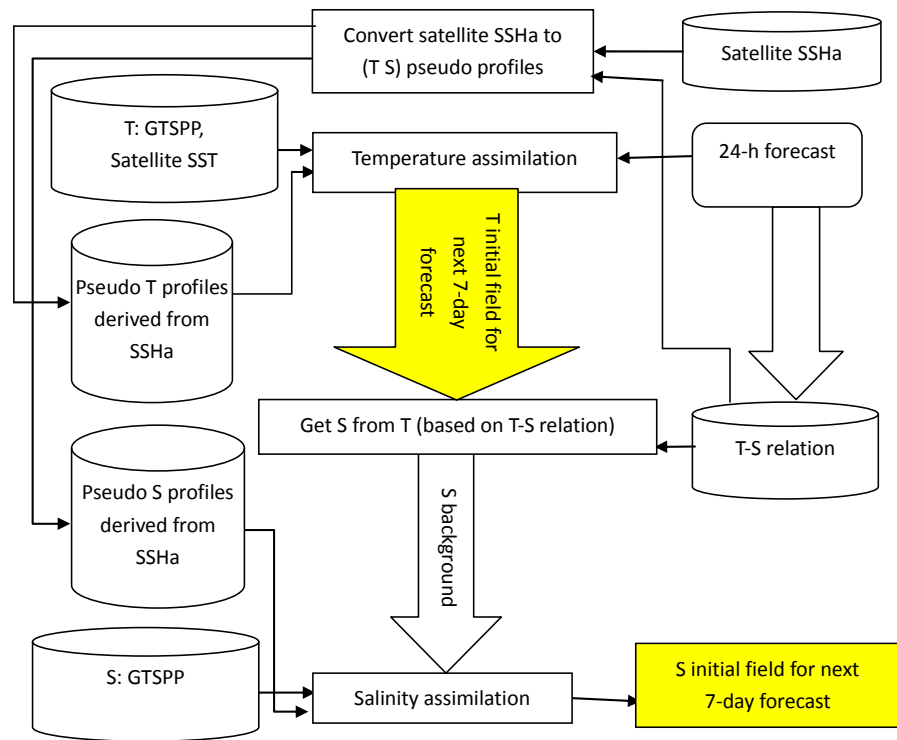


Fig. 3. Vertical distributions of numbers of observations for temperature (red) and salinity (blue) from conventional (a) and Argo data (b).

487



488

489 Fig. 4. Flow chart of multi-grid 3DVAR operational procedure.

490

491

492

493

494

495

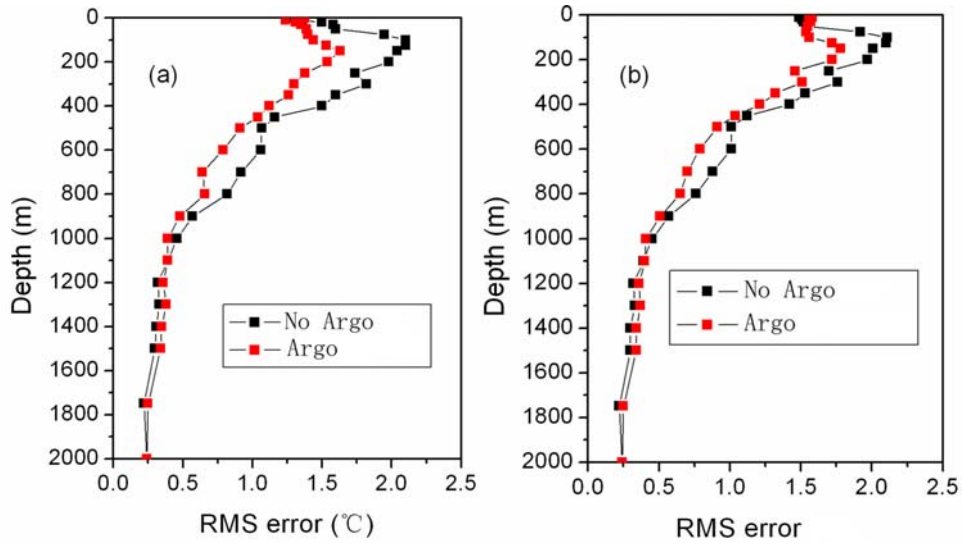
496

497

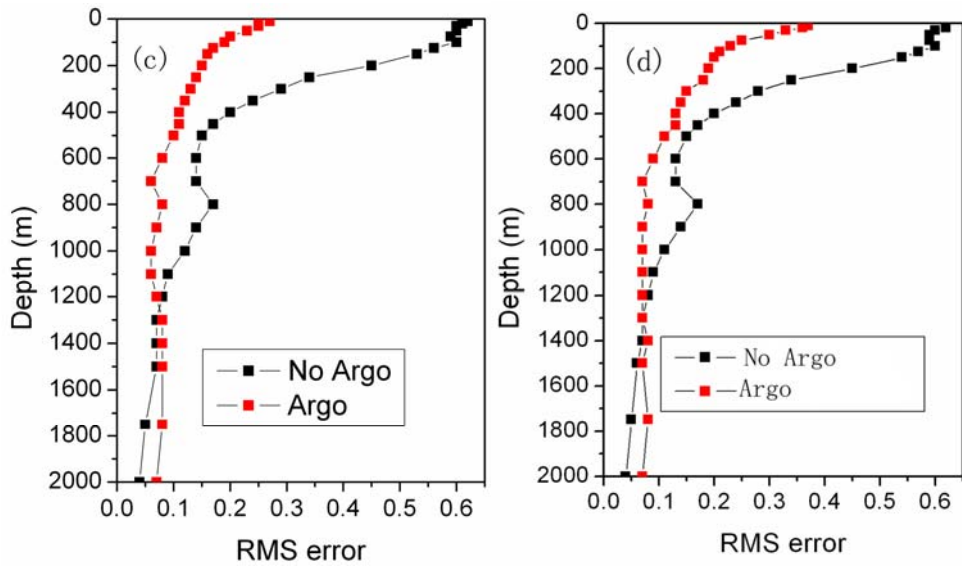
498

499

500



501



502

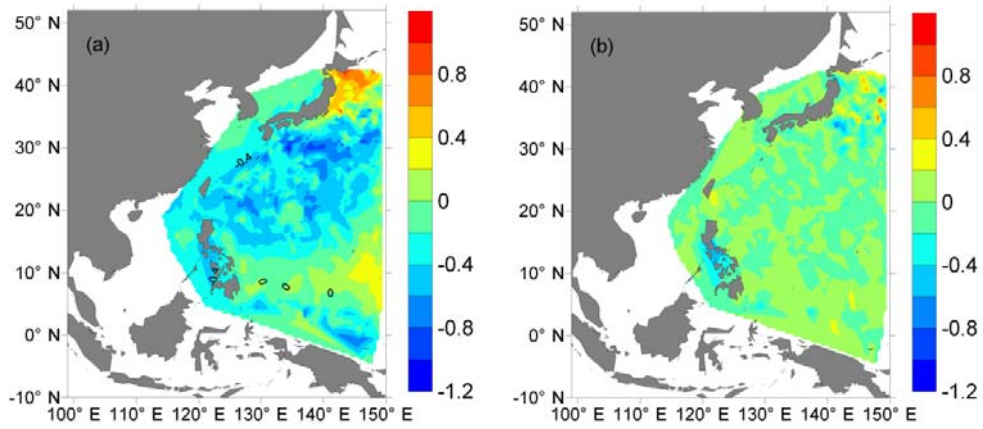
503 Fig. 5. Vertical dependence of temperature (a, b, $^{\circ}\text{C}$) and salinity (c, d, psu) H-RMSEs
 504 in 24-hour forecast (a, c) and 168-hour forecast (b, d) with and without Argo data
 505 assimilation.

506

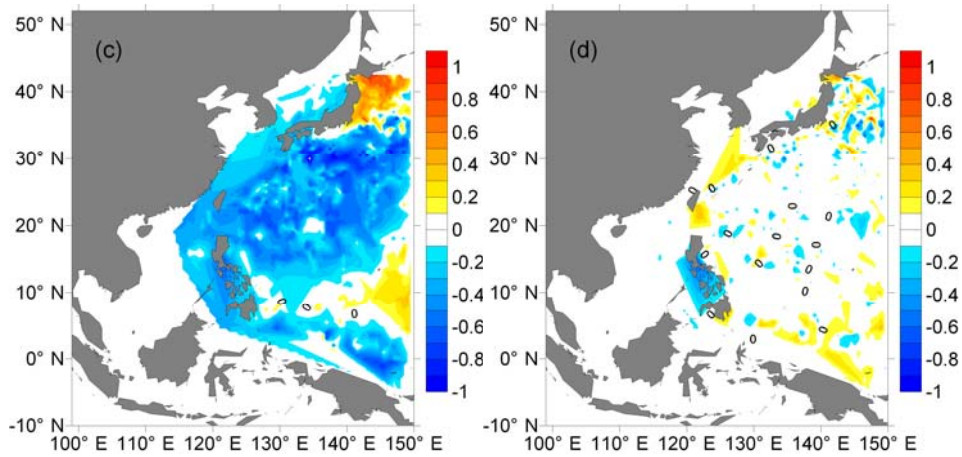
507

508

509



510



511

512 **Fig. 6.** Horizontal distribution of vertically (100-300 m) averaged temperature (a, b,
 513 °C) and salinity (c, d, psu) prediction errors in 24-hour forecast without Argo profiles
 514 assimilation(a, c) and with Argo profiles assimilation(b, d).

515

516

517

518

519

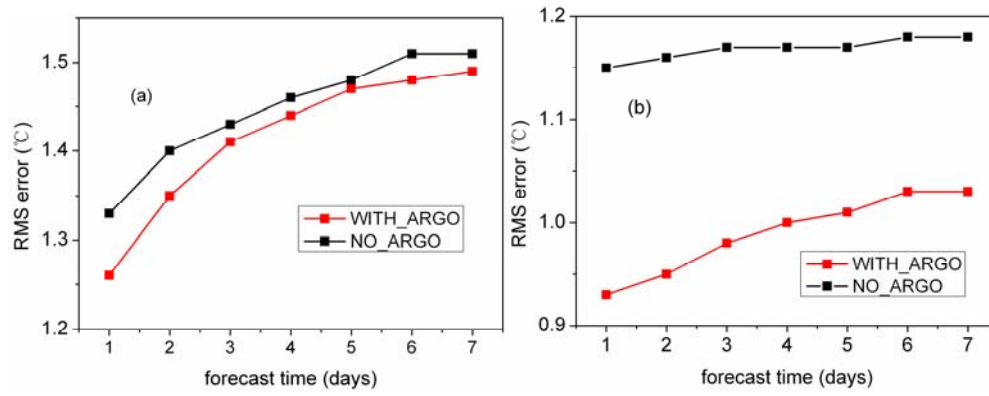


Fig. 7. Temporal variation of [temperature S-RMSEs](#) (°C) for the layers of 0-50m(a) and 50-1000m(b) in 24-hour forecast with and without Argo data assimilation.

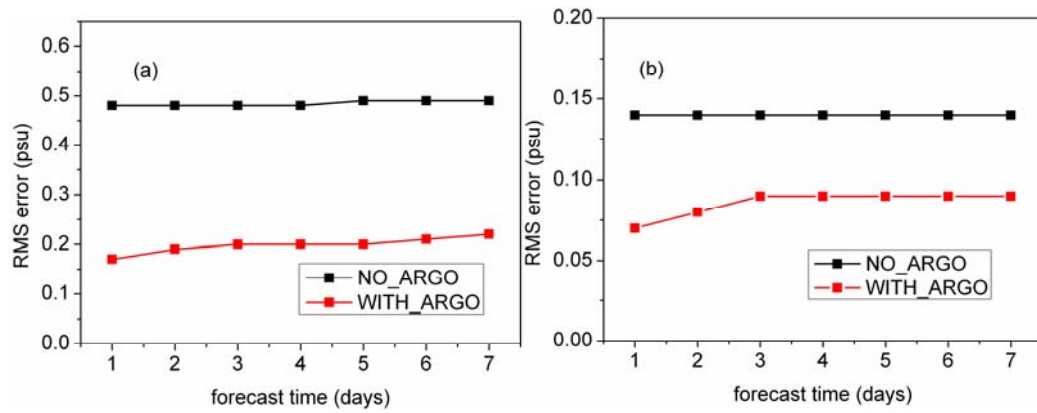


Fig. 8. Temporal variation of salinity S-RMSEs (psu) for the layers of 0-300m(a) and 300-1000m(b) in 24-hour forecast with and without Argo data assimilation.

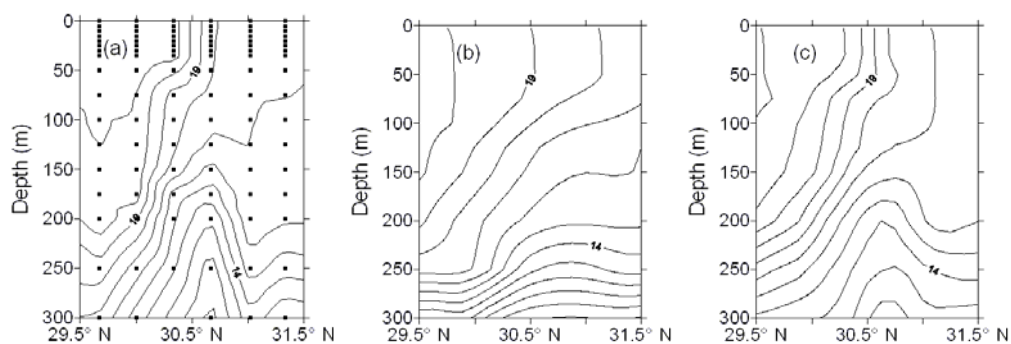


Fig. 9. Vertical temperature cross-section along 129°E south of Japan on 23 February 2008: (a) observation (dark dots: stations), (b) 24-hour forecast without assimilating Argo profiles, and (c) 24-hour forecast with assimilating Argo profiles.

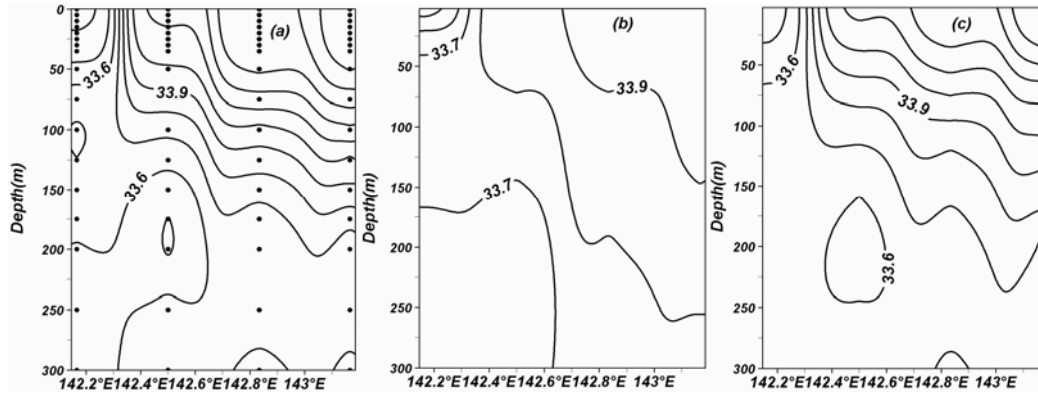


Fig. 10. Vertical salinity cross-section along 38.5°N east of Japan on 8 May 2008: (a) observation (dark dots: stations), (b) 24-hour forecast without assimilating Argo profiles, and (c) 24-hour forecast with assimilating Argo profiles.